

Reviewer: Mengxin Zhang
Date: Dec.10, 2023
Report Title: Hotel Booking Analysis and Forecasting
Author: Minghua Zhu

Peer Review for Research Project

Overall GitHub structure

1. Well organized: The GitHub repository has a clear and logical structure. This makes it easy for anyone looking at the project to find what they need quickly.
2. Data management streamlined: All data is kept in one place in the Data folder. This setup makes data management easy. The author also uploads the unzipped raw dataset from the resources, which is convenient for the viewer to access.
3. Consistent and clear file naming: Each file is named concisely and clearly, and the weekly progress documents are well labeled. However, some files have naming inconsistencies with other files in the same folder. This should be addressed for consistency.

Code / Jupyter Notebook & Output

4. Code Structure: The code is generally well structured.
5. Data Cleaning: Implements basic cleaning, but exploring advanced cleaning techniques could further enhance the data quality.
6. Visualizations: The project makes effective use of Matplotlib and Seaborn for data visualization, which greatly aids in understanding data distribution and key trends. The histograms for each column and the pie charts for type analysis are particularly impressive for their clarity and comprehensiveness.
7. Explanation of output: The author did a good job of explaining the results of several plots, providing valuable conclusions.
8. Model building: the ROC curves and confusion matrix lack detailed explanations, which could be improved to provide a more complete understanding of the model's performance.

9. Correlation analysis: It is well done and informative to calculate and display the correlation coefficients for each variable. I think using a heatmap would greatly enhance the visualization of these correlations, making it easier to see patterns and relationships between variables at a glance. A heatmap can present this complex data in a more intuitive and visually appealing way.
10. Model selection: It's better to do more analysis on their suitability for the dataset, advantages in handling categorical and continuous data, or their ability to model complex relationship

Project report

11. Clear and Structured: The report is well-written and logically organized, with clear sections and professional use of language.
12. Feature choices: It's better to explain more about why certain features were dropped. This helps to understand their impact on the model.
13. Model comparison: The part of the illustration that compares the best model to industry standards or past studies is lacking in detail. Adding more it's helpful for better performance insight.